



La Lettre de l'OCIM

Musées, Patrimoine et Culture scientifiques et techniques

122 | 2009
mars - avril 2009

Documents numériques « ramenés à la surface »

Michael Creek



Édition électronique

URL : <http://journals.openedition.org/ocim/291>

DOI : 10.4000/ocim.291

ISSN : 2108-646X

Éditeur

OCIM

Édition imprimée

Date de publication : 1 mars 2009

Pagination : 26-30

ISSN : 0994-1908

Référence électronique

Michael Creek, « Documents numériques « ramenés à la surface » », *La Lettre de l'OCIM* [En ligne], 122 | 2009, mis en ligne le 01 mars 2011, consulté le 01 mai 2019. URL : <http://journals.openedition.org/ocim/291> ; DOI : 10.4000/ocim.291

Tous droits réservés

Documents numériques « ramenés à la surface »

Michael Creek *



L'interface AquaRing facilite la recherche des documents numériques sur le monde aquatique.
© M. Creek

AquaRing est un portail sémantique d'accès aux ressources numériques d'institutions muséales travaillant autour du monde aquatique : l'un des responsables de ce projet européen présente ici les réflexions qui ont présidé à sa mise en place ainsi que les principaux objectifs qui lui ont été assignés.

Les centres de science, les musées, les aquariums, les zoos et les muséums d'Histoire naturelle sont aujourd'hui producteurs et possesseurs de nombreux documents numériques. On y trouve aussi bien des photographies et des vidéos que des articles, des collections numériques, des résultats de recherches, des logiciels ou du matériel éducatif – tous d'un grand intérêt pour le public, et venant d'institutions que le public considère comme dignes de confiance et fiables. Comment ces institutions muséales peuvent-elles travailler ensemble au niveau européen pour créer un unique point d'accès et ouvrir au public ces ressources de la façon la plus effective ?

Ce questionnement a été le point de départ du projet européen *AquaRing*, qui met en place un portail de démonstration sémantique innovateur pour les contenus numériques traitant du monde aquatique : www.aquaringweb.eu. Le contenu de cet outil est alimenté par les aquariums, les musées d'Histoire naturelle et les centres de science européens. Le but de cette ressource de démonstration en ligne, cofondée par la Commission Européenne, est d'exploiter une technologie Internet multilingue et novatrice pour faciliter l'accès intuitif.

* Michael Creek est coordonnateur de projets à Ecsite, the European Network of Science Centres and Museum.
mcreek@ecsitem.net

À la recherche des documents numériques

Le Web est aujourd'hui l'un des premiers points de référence pour toute information ; et sur des sujets scientifiques, on y trouve des documents issus d'un large éventail de sources, dont certaines sont plus scientifiquement sérieuses que d'autres. Le public dispose de plusieurs moyens de recherche pour trouver des images – il peut aller voir dans « Google Images » pour trouver des photographies de sites dans le monde entier, ou bien il peut se rendre sur des sites de réseaux sociaux comme « Flickr » où les usagers téléchargent leurs photographies qu'ils partagent avec la communauté en ligne. Mais dans ces conditions, l'internaute n'a pas de certitude sur la qualité et le sérieux scientifique de ce qu'il trouve. De nombreuses institutions donnent maintenant l'accès à leurs données sur leurs propres sites, mais ces données sont inévitablement limitées aux ressources de l'institution, et une interface de recherche bien structurée n'est pas toujours facile à mettre en place.

AquaRing propose une solution sous forme de portail : un site qui regroupe toutes ces ressources et les rend d'accès facile. L'objectif est d'assurer un contenu scientifiquement sérieux de grande qualité et de large portée en impliquant un regroupement d'aquariums réputés, de zoos, de musées d'Histoire naturelle de toute l'Europe.

Une technologie avancée pour une recherche avancée

Dès le départ, il a fallu assurer à ce portail une technologie de pointe qui, compte tenu de la rapidité des avancées technologiques, garantisse que notre outil ne serait pas rapidement dépassé. Pour AquaRing, prototype de site Internet, ce fut l'unique occasion de tester une nouvelle façon de rendre facile l'accès aux documents : l'utilisation d'une technologie sémantique d'Internet.

Sir Tim Berners-Lee, à qui l'on doit l'invention d'Internet et qui est à l'origine du « www » que nous utilisons quotidiennement, est maintenant directeur du Consortium « World Wide Web » qui supervise le développement continu d'Internet. En 2001, il a expliqué sa vision du Web sémantique dans un article du Scientific American Magazine ⁽¹⁾ :

« Aujourd'hui, la plus grosse partie du Web est conçue pour être lue par l'homme, pas pour être manipulée utilement par les programmes informatiques. Les ordinateurs manient et analysent bien les pages Web pour la mise en page et le traitement des opérations courantes –

ici un titre, là un lien vers une autre page – mais en général, les ordinateurs n'ont pas de procédé fiable pour élaborer les liens sémantiques : ceci est la page d'accueil de « Hartman and Strauss Physio Clinic », ce lien mène au curriculum vitae du Docteur Hartman. Le Web sémantique apportera une structure qui donnera du sens aux pages Web, créant un environnement là où des agents de logiciel passant de page en page peuvent effectuer facilement des tâches sophistiquées pour les utilisateurs. Un tel agent arrivant à la page Web de clinique ne saura pas seulement que la page a des mots clés tel que « traitement, médecine, physique, thérapie » (comme il est possible qu'ils soient encodés de nos jours) mais aussi que le Dr Hartman travaille à cette clinique les lundis, mercredis et vendredis et que le script prend des possibilités de dates en format année/mois/jours et renvoie des dates de rendez-vous. Et il saura tout ceci sans avoir besoin d'une intelligence artificielle à l'échelle de Hal 2001 ou C-3PO (D2R2) de la Guerre des Étoiles. À la place, ces sémantiques ont été encodées dans la page Web quand le chef de bureau de la clinique (qui n'a jamais pris Comp Sci 101) l'a façonnée utilisant des logiciels fixes pour écrire des pages Web en même temps que les ressources qui se trouvent sur le site de Physical Therapy Association ».

En d'autres termes, le Web sémantique rend les moteurs de recherche « intelligents » parce qu'il leur permet de comprendre les relations conceptuelles entre les mots. Ainsi, lorsque l'on cherche sur Google le mot « bluefin tuna », le moteur de recherche trouve les pages avec les mots « bluefin tuna »



Le « géotagging » permet un accès facile aux images grâce à l'outil de recherche géographique.

© M. Creek

écrit sur la page, ou qui se rapporte à la page. Google ne sait pas reconnaître que « bluefin tuna » est la même chose que *Thunnus thynnus*, ou qu'en Hollandais il s'agit de « blauwvintonijn ». Google ne sait pas qu'on peut le trouver des deux côtés de l'Atlantique, ou que les populations de ces espèces souffrent de sur-pêche. Bien sûr, un homme peut rechercher cette information en utilisant Google, mais dans des moteurs de recherche de Web sémantique, ce genre d'information peut être intégré dans le moteur de recherche lui-même.

Tagging : cataloguer la recherche

Ce processus de construction implique le marquage (tagging) du contenu recherché. Ces marqueurs sont connus comme métadonnées – données concernant les données. Cette méthode de classification de documents existe déjà. Mais dans le cas du Web sémantique ce sont les marqueurs des métadonnées qui sont extraits des systèmes de classifications liés entre eux. Pour une thématique telle que celle du monde aquatique, un éventail de systèmes de classification, ou ontologie, est nécessaire. *AquaRing* se sert de ASFA (Aquatic Sciences and Fisheries Abstracts), qui couvre la totalité de la documentation en matière de science, technologie, gestion, et la conservation de ressources et environnements maritimes, eau saumâtre, et eau douce, y compris leurs aspects socio-économiques et judiciaires. *AquaRing* utilise également la classification EUNIS (European University Information Systems) pour les habitats et ASFIS (Aquatic Sciences and Fisheries Information Systems) qui stocke des informations relatives aux Groupes, Ordres, Familles d'environ 11 000 espèces en relation avec la pêche et l'aquaculture. Plusieurs autres ontologies y sont annexées, afin de rendre les métadonnées aussi utiles que possible pour décrire chaque document. Puis intervient l'étape cruciale qui est au cœur du processus de ce Web sémantique – les ontologies sont liées les unes aux autres. Par exemple, nous pouvons rattacher une espèce avec son habitat, ou avec les problèmes environnementaux qui la concernent.

Les avantages du tagging

Ainsi, lorsqu'un utilisateur fait une recherche dans *AquaRing* pour « bluefin tuna » le moteur de recherche peut immédiatement trouver du matériel additionnel en relation non seulement avec l'espèce elle-même, mais aussi avec sa place dans un écosystème et la façon dont, par exemple, elle est pêchée et

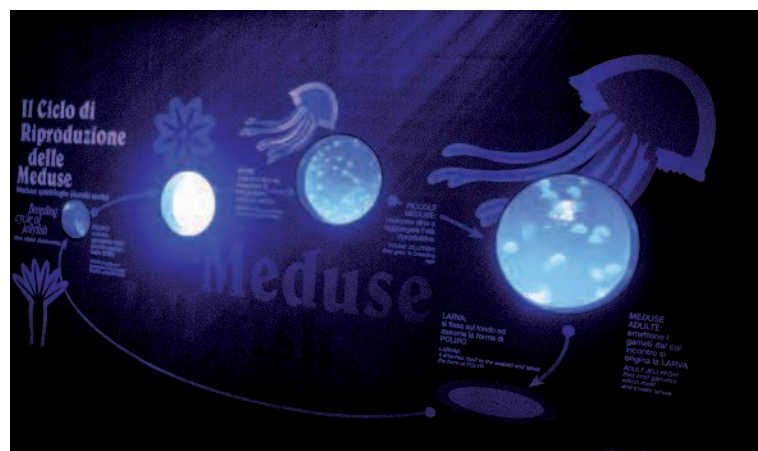
consommée. Cette technologie met les intérêts de l'utilisateur au cœur du processus.

De plus, le Web sémantique permet au moteur de recherche d'être véritablement multilingue. Ceci est réalisé par la traduction des ontologies, permettant aux documents d'être classés et reliés les uns aux autres, dans toutes les langues, dans le même système. Ainsi, quelle que soit la langue utilisée pour la recherche, le moteur de recherche est capable de récupérer les résultats appropriés qui ont été marqués dans n'importe quelle langue.

Cependant, l'avantage principal décrit par Sir Tim Berners-Lee est que le contenu devient accessible non seulement pour les utilisateurs d'Internet mais aussi pour les ordinateurs puisque les relations à l'intérieur du contenu marqué sémantiquement sont inhérentes au système de classification, ce qui ouvre des possibilités sans fin pour des extensions du système existant. On pourrait imaginer des équipements permettant d'utiliser le moteur de recherche de multiples façons – il serait ainsi possible d'utiliser un portable au restaurant pour vérifier si une variété particulière de poisson est pêchée en respect du renouvellement de l'espèce, ou vérifier sur le GPS d'une voiture quelle espèce de baleine est-on susceptible d'apercevoir en conduisant le long de la côte.

L'avenir du Web sémantique

Bien sûr, le succès du Web sémantique dépend de sa mise en application. La position dominante de Google sur le marché des moteurs de recherche est indiscutable ; cependant des évolutions récentes indiquent que le Web sémantique pourrait être un bon investissement pour ses rivaux. En mars 2008, Yahoo a



De nouvelles images sont maintenant accessibles au grand public à travers *AquaRing*.
© Aquarium de Gênes

annoncé un soutien Web sémantique à la nouvelle plateforme « Yahoo Search Open Platform » ⁽²⁾, et quatre mois plus tard, Microsoft a suivi avec l'acquisition d'un moteur de recherche sémantique « Powerset » ⁽³⁾. Ce qui a provoqué des vagues en ligne – après l'immense succès de la prétendue technologie « Web 2.0 », basée autour des réseaux sociaux, du partage de contenu, des wikis et blogs. Toutefois, la question demeure : le Web 3.0 représenterait-il une avancée dans la direction de la technologie du Web sémantique ?

L'Europe souhaite imprimer sa marque sur le Web sémantique. Ainsi, en 2005 la Commission Européenne a proposé « eContentplus » ⁽⁴⁾, un programme de 4 ans sur lequel des projets comme *AquaRing* ont été fondés. Le but est de rendre le contenu numérique plus accessible, utilisable et exploitable en Europe. Le budget de 149 millions d'euros vise à s'attaquer à des barrières organisationnelles, et à faire la promotion de l'intégration des solutions techniques qui améliorent l'accessibilité et l'utilisation des matériels numériques dans un environnement multilingue. Le programme aborde des domaines de marchés spécifiques où le développement n'est pas rapide : les informations géographiques, les contenus éducatifs et les bibliothèques numériques (contenu culturel, scientifique et scolaire), en particulier la création de la Bibliothèque Numérique Européenne. Cet outil connu sous le nom de Europeana ⁽⁵⁾, a démarré en novembre 2008 et utilise pleinement les technologies du Web sémantique. Afin que les institutions puissent fournir du contenu, leurs documents doivent être marqués avec des métadonnées sémantiques.

Quel que soit le degré de succès du Web sémantique, pour un projet tel que *AquaRing*, les avantages de ce type de technologie sont nombreux. Le marquage du contenu est un moyen efficace de classification qui ouvre le contenu numérique non seulement pour un accès par des outils tels que *AquaRing* mais aussi pour toutes sortes de nouvelles interfaces potentielles. Le processus devient facile avec le système de gestion du contenu d'*AquaRing* grâce à ses caractéristiques d'annotation par collection. Plutôt que de marquer chaque document un par un, on peut simplement placer tous les documents semblables dans un classeur et marquer le dossier. Cette opération peut se faire automatiquement pendant que l'on charge des dossiers. Un seul classeur peut contenir des centaines d'éléments, accélérant ainsi le processus. Ces collections peuvent aussi être regroupées, permettant à un groupe de collections d'être marquées simultanément. Ainsi, la

façon dont les documents sont marqués suit la façon dont ils sont structurés dans les classeurs. Il en résulte une bien meilleure maîtrise des collections de documents et une façon plus efficace de les marquer.

La protection des droits d'auteur

L'autorisation légale donnée à « Creative Commons » ⁽⁶⁾, qui résout de nombreux problèmes liés au droit d'auteur, a entraîné une augmentation de la quantité de documents mis en ligne. Les institutions peuvent désormais donner l'accès à leurs images tout en gardant la propriété intellectuelle de leurs droits ; cette autorisation d'accès permet de préciser les conditions d'utilisation : indiquer la source de l'image, ne pas tirer profit de l'image, ne pas la diffuser sur d'autres supports... En conséquence, *AquaRing* a pu bénéficier de cette licence pour s'assurer que les institutions puissent fournir du contenu sans craindre une atteinte aux droits d'auteurs.



Les thèmes importants liés au développement durable, tels que la sur-pêche, bénéficient d'une visibilité importante sur *AquaRing*.

© M. Creek

AquaRing pour les institutions

Le projet *AquaRing* a été lancé avec des objectifs précis : augmenter la valeur des collections actuelles de documents numériques ; encourager le développement d'une perspective européenne dans ce domaine ; créer un espace virtuel d'apprentissage ; soutenir l'éducation formelle et informelle sur le monde aquatique.

L'aquarium de Gênes en Italie, Nausicaà, le centre national de la Mer en France, l'institut royal des Sciences naturelles de Belgique, le zoo de Rotterdam aux Pays-Bas et le musée de la Mer de Lituanie sont pour l'instant les principaux établissements fournisseurs du contenu du portail. D'autres institutions muséales sont invitées à les rejoindre. Cette démarche peut leur permettre d'accroître auprès de leurs publics les moyens d'information disponibles dans le domaine des sciences de la mer. Il s'agit aussi d'un engagement européen fort avec l'idée que ces collections ont plus de valeur ensemble que séparées et que la possibilité de profiter de la variété et la complémentarité culturelles des institutions européennes est un atout important.

Enfin, la participation au projet *AquaRing* doit donner l'opportunité aux institutions partenaires de renforcer leur visibilité – auprès du public mais également parmi les professionnels du secteur à l'échelle européenne, voire internationale.

AquaRing pour les utilisateurs

Pour les utilisateurs potentiels, les avantages sont nombreux. Au cours de l'année 2009, *AquaRing* sera accessible à toute personne cherchant des informations sur le monde aquatique. Son large éventail d'images, de vidéos, d'articles et de matériel interactif en fera une grande base de ressources pour le public. Cette ressource fournira aux professeurs et conférenciers un plus large éventail d'informations scientifiques. Le portail pourra utiliser des données complémentaires sur les environnements aquatiques, il donnera des informations essentielles pour accéder aux experts du monde aquatique. *AquaRing* sera un nouvel outil de communication présentant des informations qui intéressent les directeurs et personnels d'aquarium, de musées d'Histoire naturelle et des centres de sciences naturelles en permettant, par exemple, la réutilisation de matériaux interactifs initialement créés pour des expositions temporaires.

Notes

- (1) Scientific American, mai 2001, www.sciam.com/article.cfm?id=the-semantic-web
- (2) The Yahoo! Search Open Ecosystem, Yahoo Search Blog, 13 mars 2008
www.ysearchblog.com/archives/000527.html
- (3) Powerset rejoint Live Search, Microsoft Live Search Blog, 1^{er} juillet 2008
<http://blogs.msdn.com/livesearch/archive/2008/07/01powerset-joins-live-search.aspx>
- (4) http://ec.europa.eu/information_society/activities/econtentplus/index_en.htm
- (5) www.europeana.eu/
- (6) <http://creativecommons.org/>